

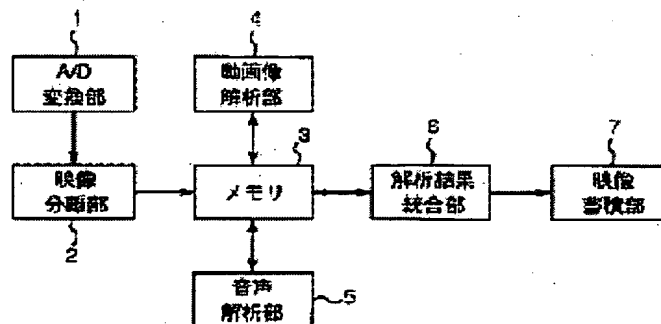
**NEWS ARTICLE CUTTING DEVICE**

**Patent number:** JP2001126050  
**Publication date:** 2001-05-11  
**Inventor:** ONO ATSUSHI; AKAGI HIROYUKI  
**Applicant:** SHARP KK  
**Classification:**  
- **International:** G06T1/00; G06T7/20  
- **European:**  
**Application number:** JP19990305380 19991027  
**Priority number(s):** JP19990305380 19991027

Report a data error here

**Abstract of JP2001126050**

**PROBLEM TO BE SOLVED:** To cut a news article from a news video. **SOLUTION:** A video separation part 2 separates an input video into a moving picture part and a sound part. A moving picture analysis part 4 analyzes the moving picture and detects a cut point picture, a telop frame and the picture of a face. A sound analysis part 5 analyzes sound and detects a silence period. An analysis result integration part 6 integrates a moving picture analysis result and a sound analysis result and cuts the news article. Thus, the news article can be cut by using the characteristics of a news video.



Data supplied from the esp@cenet database - Worldwide

**THIS PAGE BLANK (USPTO)**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-126050

(P2001-126050A)

(43) 公開日 平成13年5月11日 (2001.5.11)

(51) Int.Cl.<sup>7</sup>

識別記号

F I

テーマコード(参考)

G 0 6 T 1/00  
7/20

C 0 6 F 15/62  
15/66  
15/70

P 5 B 0 5 0  
4 7 0 A 5 B 0 5 7  
4 0 5 5 L 0 9 6  
9 A 0 0 1

審査請求 未請求 請求項の数6 O L (全 17 頁)

(21) 出願番号 特願平11-305380

(22) 出願日 平成11年10月27日 (1999. 10. 27)

(71) 出願人 000003049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 小野 敦史

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(72) 発明者 赤木 宏之

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(74) 代理人 100062144

弁理士 青山 葆 (外1名)

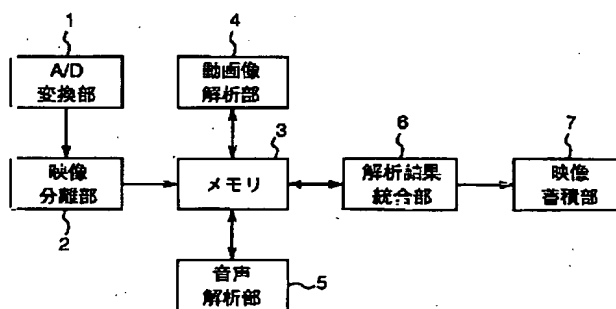
最終頁に続く

(54) 【発明の名称】 ニュース記事切り出し装置

(57) 【要約】

【課題】 ニュース映像からニュース記事を切り出す。

【解決手段】 映像分離部2は、入力映像を動画像部分と音声部分とに分離する。動画像解析部4は、上記動画像を解析してカット点画像、テロップフレーム、顔の画像を検出する。音声解析部5は、上記音声を検出して無音区間を検出する。解析結果統合部6は、上記動画像解析結果および音声解析結果を統合して、ニュース記事を切り出す。こうして、ニュース映像の特性を用いてニュース記事を切り出すことができる。



【特許請求の範囲】

【請求項1】 映像入力手段と、

上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、

上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、

上記カット点画像間の類似度を算出する類似度算出手段と、

類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴とするニュース記事切り出し装置。

【請求項2】 映像入力手段と、

上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、

上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、

上記分離された動画からテロップを検出するテロップ検出手段と、

上記テロップ検出手段によって検出されたテロップの直前に位置するカット点画像間の類似度を算出する類似度算出手段と、

上記テロップの直前に位置するカット点画像のうち類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴とするニュース記事切り出し装置。

【請求項3】 映像入力手段と、

上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、

上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、

上記分離された動画からテロップを検出するテロップ検出手段と、

上記分離された動画から顔の画像を検出する顔検出手段と、

上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度を算出する類似度算出手段と、  
上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴とするニュース記事切り出し装置。

【請求項4】 映像入力手段と、

上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、

上記映像分離手段によって分離された音声から無音部分を検出する無音検出手段と、

無音部分の間を記事として切り出す記事切り出し手段を

備えたことを特徴とするニュース記事切り出し装置。

【請求項5】 映像入力手段と、

上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、

上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、

上記分離された動画からテロップを検出するテロップ検出手段と、

上記分離された動画から顔の画像を検出する顔検出手段と、

上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度を算出する類似度算出手段と、  
上記映像分離手段によって分離された音声から無音部分を検出する無音検出手段と、

上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像を選出し、この選出カット点画像近傍に在る無音部分の間を記事として切り出す記事切り出し手段を備えたことを特徴とするニュース記事切り出し装置。

【請求項6】 請求項1乃至請求項5の何れか一つに記載のニュース記事切り出し装置において、

上記記事切り出し手段によって記事を切り出すに先立って、上記カット点画像検出手段によって検出された各カット点画像間のうち、コマーシャルメッセージに該当するカット点画像間を検出して除去するコマーシャルメッセージ除去手段を備えたことを特徴とするニュース切り出し装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、ニュース映像をデータベース化する場合に検索単位となるニュース記事を映像から自動的に切り出すニュース記事切り出し装置に関する。

【0002】

【従来の技術】従来のニュース映像データベースにおいては、情報処理学会誌Vol.37 No.9 “Informedia: CMUデジタルビデオライブラリプロジェクト”等に記載された技術によって、映像のセグメンテーションや検索のための索引付けが行なわれている。

【0003】

【発明が解決しようとする課題】しかしながら、上記従来の技術においては、映像のセグメンテーションについて要素技術が列挙されてはいるものの、具体的な解は開示されていない。

【0004】そこで、この発明の目的は、ニュース映像からニュース映像の特性を用いてニュース記事を切り出すことができるニュース記事切り出し装置を提供することにある。

## 【0005】

【課題を解決するための手段】上記目的を達成するため、第1の発明のニュース記事切り出し装置は、映像入力手段と、上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、上記カット点画像間の類似度を算出する類似度算出手段と、類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴としている。

【0006】上記構成によれば、映像入力手段によって入力されたニュース映像が、映像分離手段によって動画部分と音声部分とに分離される。そして、カット点画像検出手段によって、上記分離された動画から、動画の変化点であるカット点画像が検出される。さらに、類似度算出手段によって、上記カット点画像間の類似度が算出される。そうすると、記事切り出し手段によって、類似度の高いカット点画像の間が記事として切り出される。

【0007】また、第2の発明のニュース記事切り出し装置は、映像入力手段と、上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、上記分離された動画からテロップを検出するテロップ検出手段と、上記テロップ検出手段によって検出されたテロップの直前に位置するカット点画像間の類似度を算出する類似度算出手段と、上記テロップの直前に位置するカット点画像のうち類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴としている。

【0008】上記構成によれば、映像入力手段によって入力されたニュース映像が、映像分離手段によって動画部分と音声部分とに分離される。そして、カット点画像検出手段によって、上記分離された動画から、動画の変化点であるカット点画像が検出される。また、テロップ検出手段によって、上記分離された動画からテロップが検出される。さらに、類似度算出手段によって、上記検出されたテロップの直前に位置するカット点画像間の類似度が算出される。そうすると、記事切り出し手段によって、上記テロップの直前に位置するカット点画像のうち類似度の高いカット点画像の間が記事として切り出される。

【0009】また、第3の発明のニュース記事切り出し装置は、映像入力手段と、上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、上記分離された動画からテロップを検出するテロップ検出手段と、上記分離された動画から

顔の画像を検出する顔検出手段と、上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度を算出する類似度算出手段と、上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像の間を記事として切り出す記事切り出し手段を備えたことを特徴としている。

【0010】上記構成によれば、映像入力手段によって入力されたニュース映像が、映像分離手段によって動画部分と音声部分とに分離される。そして、カット点画像検出手段によって、上記分離された動画から、動画の変化点であるカット点画像が検出される。また、テロップ検出手段によって、上記分離された動画からテロップが検出される。また、顔検出手段によって、上記分離された動画から顔の画像が検出される。さらに、類似度算出手段によって、上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度が算出される。そうすると、記事切り出し手段によって、上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像の間が記事として切り出される。

【0011】また、第4の発明のニュース記事切り出し装置は、映像入力手段と、上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、上記映像分離手段によって分離された音声から無音部分を検出する無音検出手段と、無音部分の間を記事として切り出す記事切り出し手段を備えたことを特徴としている。

【0012】上記構成によれば、映像入力手段によって入力されたニュース映像が、映像分離手段によって動画部分と音声部分とに分離される。そして、無音検出手段によって、上記分離された音声から無音部分が検出される。そうすると、記事切り出し手段によって、上記検出された無音部分の間が記事として切り出される。

【0013】また、第5の発明のニュース記事切り出し装置は、映像入力手段と、上記映像入力手段によって入力されたニュース映像を音声と動画とに分離する映像分離手段と、上記映像分離手段によって分離された動画から、動画の変化点であるカット点画像を検出するカット点画像検出手段と、上記分離された動画からテロップを検出するテロップ検出手段と、上記分離された動画から顔の画像を検出する顔検出手段と、上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度を算出する類似度算出手段と、上記映像分離手段によって分離された音声から無音部分を検出する無音検出手段と、上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像を選出し、この選出カット点画像近傍に在る無音部分の間を記事として切り出す記事切り出し手段を備えたことを特徴

としている。

【0014】上記構成によれば、映像入力手段によって入力されたニュース映像が、映像分離手段によって動画部分と音声部分に分離される。そして、カット点画像検出手段によって、上記分離された動画から、動画の変化点であるカット点画像が検出される。また、テロップ検出手段によって、上記分離された動画からテロップが検出される。また、顔検出手段によって、上記分離された動画から顔の画像が検出される。また、無音検出手段によって、上記分離された音声から無音部分が検出される。さらに、類似度算出手段によって、上記カット点画像のうち、上記テロップ検出手段によって検出されたテロップの直前に位置して顔が映っているカット点画像間の類似度が算出される。そうすると、記事切り出し手段によって、上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像が選出され、この選出カット点画像近傍に在る無音部分の間が記事として切り出される。

【0015】また、上記第1乃至第5の発明のニュース記事切り出し装置は、上記記事切り出し手段によって記事を切り出すに先立って、上記カット点画像検出手段によって検出された各カット点画像間のうち、コマーシャルメッセージ(CM)に該当するカット点画像間を検出して除去するCM除去手段を備えることが望ましい。

【0016】上記構成によれば、ニュース記事が切り出されるに先立って、CM除去手段によって、上記検出された各カット点画像間のうち、CMに該当するカット点画像間が検出されて除去される。したがって、以後に行われる上記記事切り出し手段による記事切り出しの際には、上記CMの区間は除外されてニュース記事のみが切り出される。

【0017】

【発明の実施の形態】以下、この発明を図示の実施の形態により詳細に説明する。図1は、本実施の形態のニュース記事切り出し装置のブロック図である。

【0018】A/D変換部1は、入力された映像をA/D変換してデジタル化する。映像分離部2は、A/D変換部1によってデジタル化された映像を動画部分と音声部分に分離する。こうして分離された動画データと音声データはメモリ3に保存される。こうしてメモリ3に保存された動画は、動画解析部4によって解析される。また、メモリ3に保存された音声は、音声解析部5によって解析される。そして、動画および音声の夫々の解析結果はメモリ3に格納される。

$$D = \sum_{i=1}^N \text{abs}(\text{hist1}[i] - \text{hist2}[i]) \quad \dots (1)$$

但し、i：ヒストグラムの量子化レベル番号(最大値N)

ステップS18で、差分Dが閾値THより小さいか否かが判別される。その結果、上記閾値THより小さければステップS19に進み、閾値TH以上であればステップS20

【0019】解析結果統合部6は、上記動画解析部4および音声解析部5による解析結果をメモリ3から読み出して統合し、後に詳述するようにしてニュース記事を切り出す。こうして切り出されたニュース記事は、映像蓄積部7に蓄積されるのである。

【0020】図2は、図1に示すニュース記事切り出し装置によって実行されるニュース記事切り出し手順の概略を示すフローチャートである。以下、図2に従って、ニュース記事切り出し手順について説明する。

【0021】まず、ステップS1で、上記映像分離部2によって、A/D変換部1からのデジタル映像データが動画データと音声データとに分離される。ステップS2で、動画解析部4によって、上記分離された動画データに基づいて動画が解析される。尚、この動画解析によって、後に詳述するように、動画の変化点であるカット点画像や、テロップが映っているフレーム(テロップフレーム)や、人物の顔が映っているフレームが検出される。

【0022】ステップS3で、上記音声解析部5によって、上記分離された音声データに基づいて音声解析される。尚、この音声解析によって、後に詳述するように、無音区間が検出される。ステップS4で、解析結果統合部6によって、動画解析結果と音声解析結果とが統合されてニュース記事の切り出しが行われる。

【0023】図3は、図2に示すニュース記事切り出し手順におけるステップS2の動画解析時に行われるカット点画像検出処理動作のフローチャートである。先ず、ステップS11で、フレーム数frameが0に初期化される。ステップS12で、直前フレームの色相ヒストグラムhist0および現フレームの色相ヒストグラムhist1が0に初期化される。ステップS13で、上記メモリ3に格納されている動画データから処理すべきフレームデータが読み出される。ステップS14で、色相ヒストグラムhist1が生成されて更新される。尚、色相ヒストグラムhist1の生成については後に詳述する。ステップS15で、フレーム数frameがインクリメントされる。

【0024】ステップS16で、frame=1であるか否かが判別される。その結果、frame=1であればステップS21に進む一方、そうでなければステップS17に進む。ステップS17で、直前フレームの色相ヒストグラムhist2と現フレームの色相ヒストグラムhist1との差分Dが求められる。尚、差分Dは式(1)によって算出するが、他の計算方法によって算出しても構わない。

に進む。ステップS19で、直前フレームからの色相ヒストグラムの変化量が小さいために現フレームは動画の変化点とは見なされず、非カット点画像であると判定され

る。ステップS20で、現フレームはカット点画像であると判定される。そして、例えばカット点画像の位置を表わすカット点画像テーブルに登録される。ステップS21で、直前フレームの色相ヒストグラムhist2が、現フレームの色相ヒストグラムhist1で更新される。ステップS22で、メモリ3に未処理のフレームデータが在るか否かが判別される。その結果、在ればステップS13に戻り、次のフレームの処理に移行する。なければカット点画像検出処理動作を終了する。

【0025】このように、本実施の形態においては、現フレームの色相ヒストグラムhist1が直前フレームの色

$$H = \cos^{-1} \left( \frac{(G-B) + (G-R)}{2\sqrt{(G-B)^2 - (G-R)(B-R)}} \right) \quad \dots (2)$$

$$S = \frac{\text{Max}(R, G, B) - \text{Min}(R, G, B)}{V}$$

$$V = \text{Max}(R, G, B)$$

ステップS33で、式(3)によってヒストグラムのインク  
 $\text{hist}[i] = \text{hist}[i] + 1$

$i = H/H\_QUANT$

但し、H\_QUANT：色相の量子化定数

ステップS34で、当該フレームデータに未処理画素が在るか否かが判別される。その結果、在ればステップS31に戻って次の画素値R, G, Bの処理に移行する。なければ色相ヒストグラム生成処理動作を終了する。

【0027】図5は、図2示すニュース記事切り出し手順におけるステップS2の動画像解析時に行われるテロップフレーム検出処理動作のフローチャートである。以下の説明においては、横書きのテロップに関する検出方法を例に説明するが、縦書きのテロップを検出する場合にもx軸とy軸とを入れ換えれば同様に実行できる。

【0028】ステップS41で、フレーム数frameが0に初期化される。ステップS42で、直前フレームのエッジ画像edge2および現フレームのエッジ画像edge1が0に初期化される。ステップS43で、メモリ3に格納されている動画像データから処理すべきフレームデータが読み出される。ステップS44で、エッジ画像edge1が生成されて更新される。尚、エッジ画像の生成については後に詳述する。ステップS45で、フレーム数frameがインクリメントされる。

【0029】ステップS46で、frame=1であるか否かが判別される。その結果、frame=1であればステップS55に進む一方、そうでなければステップS47に進む。ステップS47で、後述する投影ヒストグラム生成方法によって、エッジ画像edge1のy軸への投影ヒストグラムが生成される。ステップS48で、上記ステップS47において生成されたヒストグラムが解析されて、テロップの候補領域となる山の範囲[y1, y2]が閾値等に基づいて検出される。ここで、通常、テロップの周囲にはエッジが集中している。そのために、横書きの場合には、図6

相ヒストグラムhist2に対して閾値TH以上変化した場合には、現フレームは動画の変化点であると見なし、現フレームをカット点画像として検出するのである。

【0026】図4は、図3示すカット点画像検出処理動作の上記ステップS14において実行される色相ヒストグラム生成処理動作のフローチャートである。まず、ステップS31で、図3示すカット点画像検出処理動作の上記ステップS13においてメモリ3から取り込まれたフレームデータから、1画素の画素値R, G, Bが読み出される。ステップS32で、式(2)によって座標変換が行われる。

リメントが行なわれる。

… (3)

に示すようなy軸への投影ヒストグラムには山が検出される。そこで、上記ステップS48においては、y軸への投影ヒストグラムの山を検出してテロップの候補領域とするのである。次に、ステップS49で、上記ステップS48における山の範囲の検出結果に基づいて、山が在るか否かが判別される。その結果、山が在ればテロップの候補領域は在りとしてステップS50に進む一方、山がなければテロップの候補領域は無しとして上記ステップS55に進む。

【0030】ステップS50で、上記y1からy2までの範囲のエッジがx軸に投影されてエッジ画像の投影ヒストグラムが生成される。ステップS51で、上記ステップS50において生成されたヒストグラムから、文字部分の山の範囲が閾値等に基づいて検出される。ステップS52で、上記ステップS51における山の範囲の検出結果に基づいて、山が在るか否かが判別される。その結果、山が在ればステップS54に進み、なければステップS53に進む。

【0031】ステップS53で、現フレームが非テロップフレームであると判定される。ステップS54で、現フレームがテロップフレームであると判定される。そして、例えばテロップフレームの位置を表わすテロップフレームテーブルに登録される。ステップS55で、直前フレームのエッジ画像edge2が現フレームのエッジ画像edge1で更新される。ステップS56で、メモリ3に未処理のフレームデータが在るか否かが判別される。その結果、在ればステップS43に戻り、次のフレームの処理に移行する。なければテロップフレーム検出処理動作を終了する。

【0032】このように、本実施の形態においては、生成した上記エッジ画像edge1のy軸への投影ヒストグラ

ムに山が在り、且つ、x軸への投影ヒストグラムにも山が在る場合には、現フレームにテロップ文字列が在ると判定し、現フレームをテロップフレームとして検出するのである。

【0033】図7は、図5示すテロップフレーム検出処理動作の上記ステップS44において実行されるエッジ画像生成処理動作のフローチャートである。ステップS61で、現フレームのエッジ画像edge1が0に初期化される。ステップS62で、図5示すテロップフレーム検出処理動作の上記ステップS43においてメモリ3から取り込

$$\begin{aligned} h\_edge[i][j] &= \text{abs}(\text{gray}[i-1][j] - \text{gray}[i+1][j]) \quad \dots (4) \\ v\_edge[i][j] &= \text{abs}(\text{gray}[i][j-1] - \text{gray}[i][j+1]) \end{aligned}$$

ここで、gray[i][j]は、濃淡画像grayにおける座標(j, i)の画素値である。

【0034】ステップS67で、変数jの内容がインクリメントされる。ステップS68で、 $j < (W-1)$ であるか否かが判別される。その結果、 $j < (W-1)$ であればステップS66に戻って水平エッジ及び垂直エッジの算出が継続される。一方、 $j \geq (W-1)$ であればステップS69に進む。ステップS69で、変数iがインクリメントされる。ステップS70で、 $i < (H-1)$ であるか否かが判別される。その結果 $i < (H-1)$ であればステップS65に戻って水平エッジおよび垂直エッジの算出が継続される。一方、 $i \geq (H-1)$ であればエッジ画像生成処理動作を終了する。

【0035】つまり、本実施の形態においては、 $1 \leq j \leq (W-1)$ 及び $1 \leq i \leq (H-1)$ の範囲で求めた水平方向の両隣画素のV値の差の絶対値である $h\_edge[i][j]$ を画素値とする画像 $h\_edge$ と、 $1 \leq j \leq (W-1)$ 及び $1 \leq i \leq (H-1)$ の範囲で求めた垂直方向の両隣画素のV値の差の絶対値である $v\_edge[i][j]$ を画素値とする画像 $v\_edge$ とをもって、上記エッジ画像edgeとするのである。

【0036】尚、本実施の形態においては、上述の方法によってエッジ画像edgeを生成するのであるが、それに限定されるのではなく他のエッジ検出方法を用いても差し支えない。

$$\begin{aligned} xhist[j] &= xhist[j] + \text{Min}(h\_edge1[i][j], h\_edge2[i][j]) \quad \dots (5) \\ yhist[j] &= yhist[j] + \text{Min}(v\_edge1[i][j], v\_edge2[i][j]) \end{aligned}$$

但し、本処理動作が、図5示すテロップフレーム検出処理動作の上記ステップS47において呼び出された場合には、y軸への投影ヒストグラムyhistが算出される。一方、テロップフレーム検出処理動作の上記ステップS50から呼び出された場合には、x軸への投影ヒストグラムxhistが算出される。ステップS75で、未処理画素が在るか否かが判別される。その結果、在ればステップS74に戻って次の画素に関する処理に移行し、なければ投影ヒストグラム生成処理動作を終了する。

【0039】図9は、図2に示すニュース記事切り出し手順におけるステップS2の動画像解析時に行われる人物の顔検出処理動作のフローチャートである。尚、本実

施の形態においては、図10に示す状態遷移モデルと呼ばれる階層構造を有するモデルの照合によって顔検出を行っているが、ニューラルネットワークやその他の手法を用いても差し支えない。

【0037】図8は、図5示すテロップフレーム検出処理動作の上記ステップS47あるいはステップS50において実行される投影ヒストグラム生成処理動作のフローチャートである。ステップS71で、図5示すテロップフレーム検出処理動作の上記ステップS44において、図7に示すエッジ画像生成処理動作に従って生成された現フレームのエッジ画像edge1(h\_edge1, v\_edge1)、および、作業バッファ等に保持されている前フレームのエッジ画像edge2(h\_edge2, v\_edge2)が入力される。

【0038】ステップS72で、投影する範囲(xmin, ymin)～(xmax, ymax)が設定される。但し、本処理動作が図5示すテロップフレーム検出処理動作の上記ステップS47において呼び出された場合には、エッジ画像edge1, edge1の全体が対象となるために、投影範囲は(0, 0)～(W-1, H-1)となる。また、テロップフレーム検出処理動作の上記ステップS50から呼び出された場合には、投影範囲は(0, y1)～(W-1, y2)となる。ステップS73で、y軸への投影ヒストグラムyhistおよびx軸への投影ヒストグラムxhistが0に初期化される。ステップS74で、上記ステップS72において設定された投影範囲内の一つの画素に関して、y軸への投影ヒストグラムyhistおよびx軸への投影ヒストグラムxhistが式(5)によって生成される。

【0040】ステップS81で、上記メモリ3から顔の検出用の画像が入力される。ステップS82で、上記入力された画像が、隣接する画素が類似色であるような画素の集合でなる領域に分割される。ステップS83で、上記分割された各領域の色、位置、形状の特徴量が抽出される。ステップS84で、上記各領域(領域数N)が、図10に示す状態遷移モデルの初期状態であるcolor\_segなる状態ラベルが与えられることによって初期化される。ステッ



プS85で、領域番号*i*と状態が変化した領域数を表す変数changeとの夫々が、0に初期化される。

【0041】ステップS86で、領域*i*の特徴量と、領域*i*が遷移可能な状態への遷移する場合に満たすべき状態遷移ルールとの照合が行なわれる。その結果、領域*i*が如何なる状態遷移ルールをも満たさない場合にはステップS88に進む。一方、満たす場合にはステップS87に進む。ステップS87で、領域*i*の状態ラベルが、満たしている状態遷移ルールに対応する状態の状態ラベルに更新される。そうした後、変数changeの内容がインクリメントされる。例えば、領域*i*の状態ラベルがcolor\_segであり、図10に示す状態遷移モデルを用いる場合を考えると、状態ラベルcolor\_segから遷移可能な状態はskin\_segおよびblack\_segである。この場合、領域*i*が上記両状態に遷移するために満たすべき状態遷移ルールは、図10において上記状態ラベルcolor\_segから状態ラベルskin\_segおよび状態ラベルblack\_segへの矢印に設定されているIsSkin、及びIsBlackである。すなわち、領域*i*の特徴量が状態遷移ルールIsSkinを満たしていれば領域*i*の状態ラベルをskin\_segに更新する。同様に、状態遷移ルールIsBlackを満たしていればblack\_segに更新するのである。

【0042】ステップS88で、領域番号*i*がインクリメントされる。ステップS89で、領域番号*i*が領域数*N*より小さいか否かが判別される。その結果、*i* < *N*であればステップS86に戻って次の領域に対する処理に移行する。一方、*i* ≥ *N*であればステップS90に進む。ステップS90で、change = 0であるか否か、つまり状態が遷移した領域が在るか否かが判別される。その結果、在ればステップS85に戻る。こうして、上述の処理が、状態ラベルが変化した領域が存在しなくなるまで繰り返される。

【0043】ステップS91で、総ての領域の状態ラベルをチェックすることによって、状態ラベルfaceを持つ領域が存在するか否かが判別される。その結果、存在すればステップS92に進み、存在しなければステップS93に進む。ステップS92で、人物の顔が検出されたとして、例えば人物の顔があるフレームの位置を表わす顔フレームテーブルに登録される。そうした後、人物の顔検出処理動作を終了する。ステップS93で、人物の顔は検出されなかったとして、人物の顔検出処理動作を終了する。

【0044】このように、本実施の形態においては、入力画像を類似色の領域に分割し、各領域の特徴量を抽出し、各領域の特徴量が図10に示す状態遷移モデルの状態遷移ルールを満たしていれば当該領域の状態を遷移させ、この処理を総ての領域が状態遷移しなくなるまで繰り返す。そして、状態ラベルfaceを持つ領域が存在した場合には、人物の顔を検出したと判断するのである。

【0045】図11は、図2に示すニュース記事切り出し手順におけるステップS3の音声解析時に行われる無

音区間検出処理動作のフローチャートである。先ず、ステップS101で、無音区間であることを表す変数SilenceがFALSEに初期化される。ステップS102で、区間[sp, ep]の長さ分の音声データが読み込まれる。ステップS103で、上記読み込まれた音声データから音声パワー*p*が算出される。ステップS104で、上記音声パワー*p*の分散値が式(6)によって算出される。

$$\text{Var} = \frac{1}{N} \sum_{i=1}^N (p_i - \bar{p})^2 \quad \dots (6)$$

$$\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i$$

ステップS105で、上記算出された分散値Varが閾値THより小さいか否かが判別される。その結果、Var < THであれば区間[sp, ep]は無音区間であると判断されステップS108に進む。一方、Var ≥ THであれば無音区間ではないと判断されステップS106に進む。

【0046】ステップS106で、上記変数SilenceがTRUEであるか、つまり直前の処理区間は無音区間であるか否かが判別される。その結果、TRUEでなければ上記ステップS101に戻って、同様の処理が繰り返される。一方、TRUEであればステップS107に進む。ステップS107で、後述するようにステップS109、S111において値が設定された始端startと終端endに基づいて、無音区間[start, end]が検出される。そして、無音区間の位置を表わす無音区間テーブルに登録される。そうした後、上記ステップS101に戻って、同様の処理が繰り返される。

【0047】ステップS108で、上記変数SilenceがTRUEであるか否かが判定される。その結果、TRUEでなければ、現在の区間[sp, ep]は無音区間の開始点であるとしてステップS109に進む。一方、TRUEであれば、現在の区間[sp, ep]は直前の無音区間の継続区間であるとしてステップS111に進む。ステップS109で、無音区間の始端startにspが設定される。ステップS111で、変数SilenceにTRUEが設定される。そうした後にステップS112に進む。ステップS111で、無音区間の終端endにepが設定される。ステップS112で、未処理の音声データが在るか否かが判別される。その結果、在ればステップS102に戻って次の音声データの処理に移行する。そして、上記ステップS105においてVar ≥ THと判定され、上記ステップS106において直前の処理区間は無音区間であると判定されると、上記ステップS107において無音区間[start(=sp), end(=ep)]が検出されるのである。一方、未処理の音声データがなければ無音区間検出処理動作を終了する。

【0048】このように、本実施の形態においては、音声区間[sp, ep]におけるパワー*p*の分散値Varが閾値THより小さい場合には、区間[sp, ep]は無音区間であると判断する。さらに、直前区間が無音区間であれば区間[sp, ep]は上記直前の無音区間の継続区間であると判定

する。一方、直前区間が無音区間でなければ区間[sp, ep]は無音区間の開始点であると判定する。そして、次に分散値Varが閾値TH以上になると、無音区間[start (=sp), end(=ep)]を検出するのである。

【0049】図12は、図2示すニュース記事切り出し手順のステップS4において、解析結果統合部6によって行われるニュース記事切り出し処理動作のフローチャートである。尚、本ニュース記事切り出し処理動作においては、動画像解析部4によって図3に示すカット点画像検出処理動作に従って検出されたカット点画像に基づいて、ニュース記事を切り出すものである。

【0050】ステップS121で、図3に示すカット点画像検出処理動作によって上記カット点画像テーブルに登録されているカット点画像の集合Cut:  $\{c_i | i = 1, 2, \dots, N_{cut}\}$  が得られる。そして、この集合 $\{c_i\}$ を対象として、後述するようなクラスタリングによって、第1クラスタに属するカット点画像の集合Clst:  $\{clst_i | i = 1, 2, \dots, N_{clst}\} \subset Cut$  が得られる。ステップS122で、集合 $\{clst_i\}$ のインデックスiが1に初期化される。

【0051】ステップS123で、 $clst_i$ がニュース記事の始点として設定される。ステップS124で、iがインクリメントされる。ステップS125で、 $clst_i$ がニュース記事の終点として設定される。こうして一つのニュース記事が切り出されるのである。ステップS126で、iが最大値  $N_{clst}$  よりも小さいか否かが判別される。その結果、 $i < N_{clst}$  であれば上記ステップS123に戻って次のニュース記事の切り出し処理に移行する。一方、 $i \geq N_{clst}$  であればニュース記事切り出し処理動作を終了する。

【0052】図13は、図12に示すニュース記事切り出し処理動作のステップS121において実行されるクラスタリング処理動作のフローチャートである。ステップS131で、総てのカット点画像間の類似度Similar(i, j)が算出される。ここで、i, jは類似度を算出する2つのカット点画像の番号である。尚、本実施の形態においては、類似度Similar(i, j)として式(1)の逆数を用いるが、他の類似度を用いても構わない。ステップS132で、頻度ヒストグラムHist[i], Hist[j]が0に初期化される。ステップS133で、類似度Similar(i, j)が閾値THより大きいと判別される。その結果、 $\text{Similar}(i, j) > TH$  であればステップS134に進み、 $\text{Similar}(i, j) \leq TH$  であればステップS135に進む。ステップS134で、頻度ヒストグラムHist[i], Hist[j]がインクリメントされる。ステップS135で、未処理の類似度Similar(i, j)が在るか否かが判別される。その結果、在れば上記ステップS133に戻って、次の類似度Similar(i, j)に対する処理に移行する。

【0053】ステップS136で、上記生成された頻度ヒストグラムHist[i], Hist[j]に基づいて最大頻度位置

Maxが検出される。ステップS137で、現在のクラスタが空集合であるか否かが判別される。その結果、空集合であればステップS139に進む一方、空集合でなければステップS138に進む。ステップS138で、上記検出された最大頻度位置Maxが第1クラスタに含まれるか否かが判別される。その結果、含まれていればステップS139に進む一方、含まれていなければクラスタリング処理動作を終了する。ステップS139で、総てのSimilar(Max, j)が閾値THより大きくなるようなjが第1クラスタに追加される。ステップS140で、頻度ヒストグラムHist[i], Hist[j]からMaxが除外される。そうした後、上記ステップS136に戻って上術の処理が繰り返され、上記ステップS138において最大頻度位置Maxが第1クラスタに含まれていないと判別されるとクラスタリング処理動作を終了するのである。

【0054】一般的に、ニュース映像においては、一つのニュース記事が終了する毎に、静止しているニュースキャスタの映像に切り換え、次のニュース記事の解説等があるから次のニュース記事の映像が開始されるようになっている。つまり、各ニュース記事の間には、静止しているニュースキャスタの映像という非常に類似した動画の変化点が存在するのである。

【0055】そこで、本実施の形態においては、上述のように、上記カット点画像の集合に対して、総てのカット点画像間の類似度Similar(i, j)を算出し、この類似度Similar(i, j)が閾値THより大きい頻度を表す頻度ヒストグラムHist[i], Hist[j]の最大頻度位置Maxを含むようにクラスタリングを行う。そして、第1クラスタに属する夫々のカット点画像clst間を一つのニュース記事として切り出すのである。

【0056】図14は、図2示すニュース記事切り出し手順のステップS4における解析結果統合部6によって行われる図12とは異なるニュース記事切り出し処理動作のフローチャートである。尚、本ニュース記事切り出し処理動作においては、上記カット点画像に加えて、動画像解析部4によって図5に示すテロップフレーム検出処理動作に従って検出されたテロップフレームに基づいて、ニュース記事を切り出すのである。

【0057】ステップS141で、図3に示すカット点画像検出処理動作によって上記カット点画像テーブルに登録されているカット点画像の集合Cut:  $\{c_i | i = 1, 2, \dots, N_{cut}\}$  が得られる。さらに、図5に示すテロップフレーム検出処理動作によって上記テロップフレームテーブルに登録されているテロップフレームの集合Telop:  $\{t_i | i = 1, 2, \dots, N_{telop}\}$  が得られる。そして、テロップフレーム $t_i$ の直前のカット点画像がカット点画像の集合 $\{c_i\}$ から抽出される。ステップS142で、図13に示すクラスタリング処理動作によってクラスタリングが行われ、第1クラスタに属するカット点画像の集合Clst:  $\{clst_i | i = 1, 2, \dots, N_{clst}\} \subset Cut$  が得られる。

ステップS143で、集合  $\{clst_i\}$  のインデックス  $i$  が 1 に初期化される。

【0058】ステップS144で、 $clst_i$  がニュース記事の始点として設定される。ステップS145で、 $i$  がインクリメントされる。ステップS146で、 $clst_i$  がニュース記事の終点として設定される。こうして一つのニュース記事が切り出されるのである。ステップS147で、 $i$  が最大値  $N_{clst}$  よりも小さいか否かが判別される。その結果、 $i < N_{clst}$  であれば上記ステップS144に戻って次のニュース記事の切り出し処理に移行する。一方、 $i \geq N_{clst}$  であればニュース記事切り出し処理動作を終了する。

【0059】上述したように、ニュース映像においては、各ニュース記事の間には静止しているニュースキャストの映像、という類似映像が存在し、この映像がニュース映像と言う動画全体の中の変化点となっている。また、上記ニュースキャストの映像の直後にはテロップフレームが存在するのが常である。

【0060】そこで、本実施の形態においては、テロップフレームの直前に在るカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像  $clst$  間を一つのニュース記事として切り出すのである。

【0061】図15は、図2示すニュース記事切り出し手順のステップS4において、解析結果統合部6によって行われる図12および図14とは異なるニュース記事切り出し処理動作のフローチャートである。尚、本ニュース記事切り出し処理動作においては、上記カット点画像およびテロップフレームに加えて、動画解析部4によって図9に示す人物の顔検出処理動作に従って検出された人物の顔に基づいて、ニュース記事を切り出すものである。

【0062】ステップS151で、図3に示すカット点画像検出処理動作によって上記カット点画像テーブルに登録されているカット点画像の集合  $Cut: \{c_i | i = 1, 2, \dots, N_{cut}\}$  が得られる。さらに、図5に示すテロップフレーム検出処理動作によって上記テロップフレームテーブルに登録されているテロップフレームの集合  $Telop: \{t_i | i = 1, 2, \dots, N_{telop}\}$  が得られる。更に、図9に示す人物の顔検出処理動作によって上記顔フレームテーブルに登録されているフレームの集合  $Face: \{f_i | i = 1, 2, \dots, N_{face}\} \subset Cut$  が得られる。そして、テロップフレーム  $t_i$  の直前のカット点画像であり且つ顔が検出されたカット点画像がカット点画像の集合  $\{c_i\}$  から抽出される。

【0063】ステップS152で、図13に示すクラスタリング処理動作によってクラスタリングが行われ、第1クラスに属するカット点画像の集合  $Clst: \{clst_i | i = 1, 2, \dots, N_{clst}\} \subset Face \subset Cut$  が得られる。ステップS153で、集合  $\{clst_i\}$  のインデックス  $i$  が 1 に初

期化される。

【0064】ステップS154で、 $clst_i$  がニュース記事の始点として設定される。ステップS155で、 $i$  がインクリメントされる。ステップS156で、 $clst_i$  がニュース記事の終点として設定される。こうして一つのニュース記事が切り出されるのである。ステップS157で、 $i$  が最大値  $N_{clst}$  よりも小さいか否かが判別される。その結果、 $i < N_{clst}$  であれば上記ステップS154に戻って次のニュース記事の切り出し処理に移行する。一方、 $i \geq N_{clst}$  であればニュース記事切り出し処理動作を終了する。

【0065】上述したように、ニュース映像におけるテロップフレームの直前には静止しているニュースキャストの映像、という類似している人物の顔の映像が存在し、この映像がニュース映像と言う動画全体の中の変化点となっている。

【0066】そこで、本実施の形態においては、テロップフレームの直前に在って、且つ、人の顔が検出されたカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像  $clst$  間を一つのニュース記事として切り出すのである。

【0067】図16は、図2示すニュース記事切り出し手順のステップS4において、解析結果統合部6によって行われる図12、図14および図15とは異なるニュース記事切り出し処理動作のフローチャートである。尚、本ニュース記事切り出し処理動作においては、音声解析部5によって図11示す無音区間検出処理動作に従って検出された無音区間に基づいて、ニュース記事を切り出すものである。

【0068】ステップS161で、図11に示す無音区間検出処理動作によって上記無音区間テーブルに登録された上記無音区間の集合  $Silent: \{s_i, e_i\} | i = 1, 2, \dots, N_{silent}\}$  が得られる。そして、集合  $\{s_i, e_i\}$  のインデックス  $i$  が 1 に初期化されるのである。

【0069】ステップS162で、上記無音区間の終点  $e_i$  がニュース記事の始点として設定される。ステップS163で、 $i$  がインクリメントされる。ステップS164で、無音区間の始点  $s_i$  がニュース記事の終点として設定される。こうして一つのニュース記事が切り出されるのである。ステップS165で、 $i$  が最大値  $N_{silent}$  よりも小さいか否かが判別される。その結果、 $i < N_{silent}$  であれば上記ステップS162に戻って次のニュース記事の切り出し処理に移行する。一方、 $i \geq N_{silent}$  であればニュース記事切り出し処理動作を終了する。

【0070】上述したように、ニュース映像には静止しているニュースキャストの映像が存在するのであるが、このニュースキャストは、次のニュース記事の解説に入る前に一次的に無言状態となる。そこで、本実施の形態においては、無音区間の間を一つのニュース記事と

して切り出すのである。

【0071】図17は、図2示すニュース記事切り出し手順のステップS4において、解析結果統合部6によって行われる図12および図14～図16とは異なるニュース記事切り出し処理動作のフローチャートである。尚、本ニュース記事切り出し処理動作においては、上記カット点画像、テロップフレーム、人物の顔および無音区間に基いて、ニュース記事を切り出すものである。

【0072】ステップS171で、図3に示すカット点画像検出処理動作によって上記カット点画像テーブルに登録されているカット点画像の集合Cut:  $\{c_i | i = 1, 2, \dots, N_{cut}\}$  が得られる。さらに、図5に示すテロップフレーム検出処理動作によって上記テロップフレームテーブルに登録されているテロップフレームの集合Telop:  $\{t_i | i = 1, 2, \dots, N_{telop}\}$  が得られる。更に、図9に示す人物の顔検出処理動作によって上記顔フレームテーブルに登録されたフレームの集合Face:  $\{f_i | i = 1, 2, \dots, N_{face}\} \subset \text{Cut}$  が得られる。更に、図11に示す無音区間検出処理動作によって上記無音区間テーブルに登録された無音区間の集合Silent:  $\{(s_i, e_i) | i = 1, 2, \dots, N_{silent}\}$  が得られる。そして、テロップフレーム $t_i$ の直前のカット点画像であり且つ顔が検出されたカット点画像がカット点画像の集合 $\{c_i\}$ から抽出される。

【0073】ステップS172で、図13に示すクラスタリング処理動作によってクラスタリングが行われ、第1クラスに属するカット点画像の集合C1st:  $\{c_{1st_i} | i = 1, 2, \dots, N_{c1st}\} \subset \text{Face} \subset \text{Cut}$  が得られる。ステップS173で、集合 $\{c_{1st_i}\}$ のインデックス $i$ が1に初期化される。

【0074】ステップS174で、 $c_{1st_i}$ がニュース記事の仮の始点startとして設定される。ステップS175で、 $i$ がインクリメントされる。ステップS176で、 $c_{1st_i}$ がニュース記事の仮の終点endとして設定される。ステップS177で、仮の終点end付近に無音区間が存在するか否かが判別される。その結果、存在する場合にはステップS179に進み、存在しない場合にはステップS178に進む。ステップS178で、 $i$ が最大値 $N_{c1st}$ よりも小さいか否かが判別される。その結果、 $i < N_{c1st}$ であれば、上記ステップS175に戻って仮の終点endの更新が行われる。一方、 $i \geq N_{c1st}$ であれば、ニュース記事切り出し処理動作を終了する。

【0075】ステップS179で、仮の始点start付近に在る無音区間終点が検出されて $S$ として設定される。ステップS180で、仮の終点end付近に在る無音区間始点が検出されて $E$ として設定される。ステップS181で、区間 $[S, E]$ がニュース記事として切り出される。ステップS182で、 $i$ が最大値 $N_{c1st}$ よりも小さいか否かが判別される。その結果、 $i < N_{c1st}$ であれば上記ステップS174に戻って次のニュース記事の切り出し処理に

移行する。一方、 $i \geq N_{c1st}$ であればニュース記事切り出し処理動作を終了する。

【0076】上述したように、ニュース映像におけるテロップフレームの直前には静止しているニュースキャストの映像、という類似している人物の顔の映像が存在し、この映像がニュース映像と言う動画全体の中の変化点となっている。また、上記ニュースキャストは、次のニュース記事の解説に入る前に一次的に無言状態となる。

【0077】そこで、本実施の形態においては、テロップフレームの直前に在って、且つ、人の顔が検出されたカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像 $c_{1st}$ から一つのニュース記事の仮の始点startと仮の終点endとを求める。そして、上記仮の始点start付近の無音区間終点 $S$ と仮の終点end付近の無音区間始点 $E$ との間を一つのニュース記事として切り出すのである。

【0078】図18は、上記解析結果統合部6によって、図12、図14～図17に示すニュース記事切り出し処理動作が行われるに先立って実行されるCM区間を除去するCM除去処理動作のフローチャートである。但し、図18に示すCM除去処理動作のフローチャートは、15秒間のCMを検出除去するものである。したがって、30秒間のCMを検出除去する場合には、図18に示すCM除去処理動作のフローチャート中における数字15を30に変更すればよい。

【0079】ステップS191で、開始カット点画像番号startが0に初期化される。ステップS192で、カット点画像間累積時間intervalが0に初期化され、終了カット点画像番号endがstart+1に初期化される。ステップS193で、図3に示すカット点画像検出処理動作によって検出されたカット点画像に間して、end-1番目のカット点画像とend番目のカット点画像との間(カット点画像間 $[end-1, end]$ )の時間が取得されて、カット点画像間累積時間intervalに加算される。ステップS194で、カット点画像間累積時間intervalが15であるか否かが判別される。その結果、interval=15であればステップS195に進み、そうでなければステップS197に進む。

【0080】ステップS195で、当該カット点画像間 $[end-1, end]$ がCM区間と判定されて映像データが削除される。そして、開始カット点画像番号startがendに更新される。こうすることによって、以降のニュース記事切り出し処理動作等においては、CM区間が処理対象から外されることになる。

【0081】ステップS196で、上記開始カット点画像番号startが総カット点画像数 $N-1$ よりも小さいか否かが判別される。その結果、start $<(N-1)$ であれば、上記ステップS192に戻って次のCMの検出処理に

移行する。一方、 $start \geq (N-1)$ であればCM除去処理動作を終了する。

【0082】ステップS197で、上記開始カット点画像番号startが、総カット点画像数 $N-1$ よりも小さいか否かが判別される。その結果、 $start < (N-1)$ であればステップS198に進み、 $start \geq (N-1)$ であればCM除去処理動作を終了する。ステップS198で、上記カット点画像間累積時間intervalが15以上であるか否かが判別される。その結果、 $interval \geq 15$ である場合にはステップS200に進み、 $interval < 15$ である場合にはステップS199に進む。ステップS199で、終了カット点画像番号endがインクリメントされる。そうした後、上記ステップS193に戻って当該CMの検出処理が継続される。ステップS200で、上記開始カット点画像番号startがインクリメントされる。そうした後、上記ステップS192に戻って、次のCMの検出処理に移行する。そして、上記ステップS196、S197において、 $start \geq (N-1)$ であると判別されるとCM除去処理動作を終了する。

【0083】このように、本実施の形態においては、図3に示すカット点画像検出処理動作によって検出されたカット点画像に間して、先頭から順次カット点画像間の累積時間を取得する。そして、カット点画像間累積時間が15秒になった場合には、そのカット点画像間はCM区間であるとしてその間の映像を削除する。そうすることによって、以後のニュース記事切り出し処理等において、CM区間を処理対象から外すことができるのである。

【0084】尚、本実施の形態においては、上述の方法によってCMを検出したが、他の方法によってCM検出を行なっても一向に構わない。

【0085】上述のように、本実施の形態においては、映像分離部2によって、入力映像を動画像部分と音声部分に分離する。そして、動画像解析部4によって上記動画像が解析され、音声解析部5によって上記音声解析される。

【0086】その場合における動画像の解析は、

(1) 現フレームの色相ヒストグラムhist1が、直前フレームの色相ヒストグラムhist2に対して閾値TH以上変化した場合には、現フレームをカット点画像として検出する。

(2) 両隣画素のV値の差の絶対値を画素値とするエッジ画像edge1のy軸への投影ヒストグラムおよびx軸への投影ヒストグラムに山が在る場合には、現フレームをテロップフレームとして検出する。

(3) 入力画像を類似色の領域に分割して特徴量を抽出し、各領域の特徴量に基づいて上記状態遷移モデルにしたがって各領域の状態の遷移を繰り返す。そして、最終的に状態ラベルfaceを持つ領域が存在する場合には、人物の顔を検出したと判断する。

【0087】また、上記音声の解析は、

(4) 音声区間[sp,ep]におけるパワーpの分散値Varが閾値THより小さく、直前区間が無音区間でなければ無音区間の始端startにspを設定する一方、上記直前区間が無音区間であれば無音区間の終端endにepを設定する。そして、次に上記分散値Varが閾値TH以上になると、区間[start,end]を無音区間として検出する。

【0088】そして、上記動画像解析結果および音声解析結果に基づいて、解析結果統合部6によって、以下の方法によってニュース記事を切り出す。

【0089】(A) 総てのカット点画像間の類似度Similar(i,j)が閾値THより大きい頻度を表す頻度ヒストグラムHist[i], Hist[j]を求め、最大頻度位置Maxを含むようにクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像clst間を一つのニュース記事として切り出す。

【0090】したがって、上記色相ヒストグラムの変化点であるカット点画像(ニュースキャスタの画像)に基づいて、ニュース映像からニュース記事を切り出すことができる。

【0091】(B) テロップフレームの直前に在るカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像clst間を一つのニュース記事として切り出す。

【0092】したがって、上記カット点画像(ニュースキャスタの画像)とテロップフレームとに基づいて、ニュース映像から更に精度よくニュース記事を切り出すことができる。

【0093】(C) テロップフレームの直前に在って、且つ、顔が検出されたカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像clst間を一つのニュース記事として切り出す。

【0094】したがって、上記テロップフレームと顔が検出されたカット点画像(ニュースキャスタの画像)とに基づいて、ニュース映像から更に精度よくニュース記事を切り出すことができる。

【0095】(D) 無音区間の間を一つのニュース記事として切り出す。したがって、音声情報(ニュースキャスタの無音区間)に基づいて、ニュース映像からニュース記事を切り出すことができる。

【0096】(E) テロップフレームの直前に在って、且つ、顔が検出されたカット点画像の集合に対して、上記類似度を用いたクラスタリングを行う。そして、第1クラスに属する夫々のカット点画像clstから一つのニュース記事の仮の始点startと仮の終点endとを求める。そして、仮の始点start付近の無音区間終点Sと仮の終点end付近の無音区間始点Eとの間を一つのニュース記事として切り出す。

【0097】したがって、上記テロップフレームと顔が検出されたカット点画像(ニュースキャスタの画像)と音声情報(ニュースキャスタの無音区間)とに基づいて、ニュース映像から更に精度よくニュース記事を切り出すことができる。

【0098】さらに、上記解析結果統合部6は、上述のようなニュース記事切り出し処理を行うに先立って、上記カット点画像の列から累積時間が15秒になるカット点画像間を検索し、累積時間が15秒であるカット点画像間をCMと確定して削除する。したがって、以後のニュース記事切り出し処理等において、上記CM区間を処理対象から外すことができるのである。

【0099】尚、この発明のニュース記事切り出し装置においては、上記色相ヒストグラムの変化点であるカット点画像に基づくニュース記事切り出し方法、上記カット点画像とテロップフレームとに基づくニュース記事切り出し方法、上記テロップフレームと顔が検出されたカット点画像とに基づくニュース記事切り出し方法、音声情報(無音区間)に基づくニュース記事切り出し方法、上記テロップフレームと顔が検出されたカット点画像と音声情報(無音区間)とに基づくニュース記事切り出し方法の総てが実現可能な構成を有する必要は無い。上記各ニュース記事切り出し方法から適宜選択すればよい。

【0100】

【発明の効果】以上より明らかなように、第1の発明のニュース記事切り出し装置は、映像分離手段によってニュース映像を動画部分と音声部分とに分離し、この分離された動画からカット点画像検出手段によってカット点画像(動画の変化点)を検出し、類似度算出手段によって上記カット点画像間の類似度を算出し、記事切り出し手段によって、類似度の高いカット点画像の間を記事として切り出すので、ニュース映像を構成する類似している動画の変化点の位置でニュース記事を切り出すことができる。

【0101】すなわち、この発明によれば、一つのニュース記事が終了して、上記動画画像がニュースキャスタの映像等に切り変わった時点を的確に検出して、ニュース記事を正しく切り出すことができるのである。

【0102】また、第2の発明のニュース記事切り出し装置は、映像分離手段によってニュース映像を動画部分と音声部分とに分離し、この分離された動画からカット点画像検出手段によってカット点画像(動画の変化点)を検出し、テロップ検出手段によってテロップを検出し、類似度算出手段によって上記テロップの直前に位置するカット点画像間の類似度を算出し、記事切り出し手段によって、上記テロップの直前に位置するカット点画像のうち類似度の高いカット点画像の間を記事として切り出すので、上記ニュース映像を構成する動画の変化点の位置であって、テロップの直前に在る類似している動画の変化点の位置で、ニュース記事を切り出すことができ

る。

【0103】すなわち、この発明によれば、テロップの映像を参照して、上記動画画像がニュースキャスタの映像等に切り変わった時点をよりの確に検出して、ニュース記事を更に正しく切り出すことができるのである。

【0104】また、第3の発明のニュース記事切り出し装置は、映像分離手段によってニュース映像を動画部分と音声部分とに分離し、この分離された動画からカット点画像検出手段によってカット点画像(動画の変化点)を検出し、テロップ検出手段によってテロップを検出し、顔検出手段によって顔の画像を検出し、類似度算出手段によって上記テロップの直前に位置して顔が映っているカット点画像間の類似度を算出し、記事切り出し手段によって、上記テロップの直前に位置して顔が映っているカット点画像のうち類似度の高いカット点画像の間を記事として切り出すので、ニュース映像を構成する動画の変化点の位置であって、テロップの直前に在る類似している人物の顔が映っている動画の変化点の位置で、ニュース記事を切り出すことができる。

【0105】すなわち、この発明によれば、テロップの映像を参照して、上記動画画像が人物の映像、すなわちニュースキャスタの映像に切り変わった時点を更に的確に検出して、ニュース記事を更に正しく切り出すことができるのである。

【0106】また、第4の発明のニュース記事切り出し装置は、映像分離手段によってニュース映像を動画部分と音声部分とに分離し、この分離された音声から無音検出手段によって無音部分を検出し、記事切り出し手段によって、上記無音部分の間を記事として切り出すので、次のニュース記事の開始の前に上記ニュースキャスタが無音状態になった時点を検出して、ニュース記事を正しく切り出すことができる。

【0107】また、第5の発明のニュース記事切り出し装置は、映像分離手段によってニュース映像を動画部分と音声部分とに分離し、この分離された動画からカット点画像検出手段によってカット点画像(動画の変化点)を検出し、テロップ検出手段によってテロップを検出し、顔検出手段によって顔の画像を検出し、上記分離された音声から無音検出手段によって無音部分を検出し、類似度算出手段によって、上記テロップの直前に位置して人物の顔が映っているカット点画像間の類似度を算出し、記事切り出し手段によって、上記テロップの直前に位置して顔が映っている類似度の高いカット点画像を選出し、この選出カット点画像近傍に在る無音部分の間を記事として切り出すので、テロップの直前に在る類似している人物の顔が映っている動画の変化点の位置の近傍の無音部分間で、ニュース記事を切り出すことができる。

【0108】すなわち、この発明によれば、テロップの映像を参照して、上記動画画像が人物の映像、すなわちニュースキャスタの映像に切り変わった時点であって、且

つ、次のニュース記事の開始前に上記ニュースキャストが無言状態になった時点を的確に検出して、ニュース記事を更に正しく切り出すことができるのである。

【0109】また、上記第1乃至第5の発明のニュース記事切り出し装置は、上記記事の切り出しに先立って、上記検出された各カット点画像間のうちCMに該当するカット点画像間を検出して除去するCM除去手段を備えれば、CM区間のニュース映像を事前に除去できる。したがって、以後に行われる上記記事切り出し手段による記事切り出しの際にはニュース記事のみを切り出すことができ、誤検出を低減した精度の高いニュース記事切り出しが可能になるのである。

【図面の簡単な説明】

【図1】 この発明のニュース記事切り出し装置のブロック図である。

【図2】 図1に示すニュース記事切り出し装置によって実行されるニュース記事切り出し手順の概略を示すフローチャートである。

【図3】 図2における動画像解析時に行われるカット点画像検出処理動作のフローチャートである。

【図4】 図3に示すカット点画像検出処理動作において実行される色相ヒストグラム生成処理動作のフローチャートである。

【図5】 図2における動画像解析時に行われるテロップフレーム検出処理動作のフローチャートである。

【図6】 テロップの候補領域と投影ヒストグラムの山との関係を示す図である。

【図7】 図5示すテロップフレーム検出処理動作において実行されるエッジ画像生成処理動作のフローチャートである。

【図8】 図5示すテロップフレーム検出処理動作にお

いて実行される投影ヒストグラム生成処理動作のフローチャートである。

【図9】 図2における動画像解析時に行われる人物の顔検出処理動作のフローチャートである。

【図10】 状態遷移モデルの一例を示す図である。

【図11】 図2における音声解析時に行われる無音区間検出処理動作のフローチャートである。

【図12】 図1における解析結果統合部によって行われるニュース記事切り出し処理動作のフローチャートである。

【図13】 図12に示すニュース記事切り出し処理動作において実行されるクラスタリング処理動作のフローチャートである。

【図14】 図12とは異なるニュース記事切り出し処理動作のフローチャート図である。

【図15】 図12および図14とは異なるニュース記事切り出し処理動作のフローチャートである。

【図16】 図12、図14および図15とは異なるニュース記事切り出し処理動作のフローチャートである。

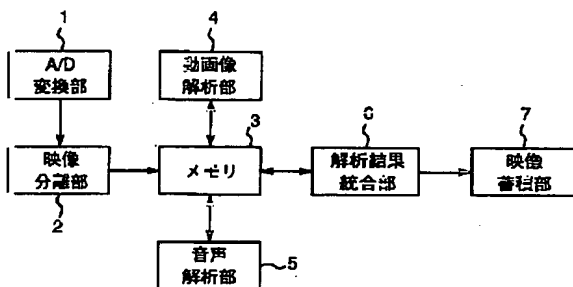
【図17】 図12および図14～図16とは異なるニュース記事切り出し処理動作のフローチャートである。

【図18】 図1における解析結果統合部によって実行されるCM除去処理動作のフローチャートである。

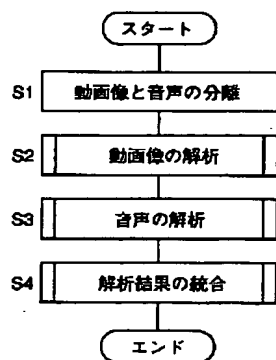
【符号の説明】

- 1…A/D変換部、
- 2…映像分離部、
- 3…メモリ、
- 4…動画像解析部、
- 5…音声解析部、
- 6…解析結果統合部、
- 7…映像蓄積部。

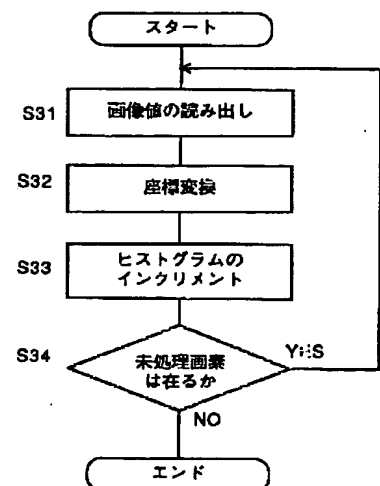
【図1】



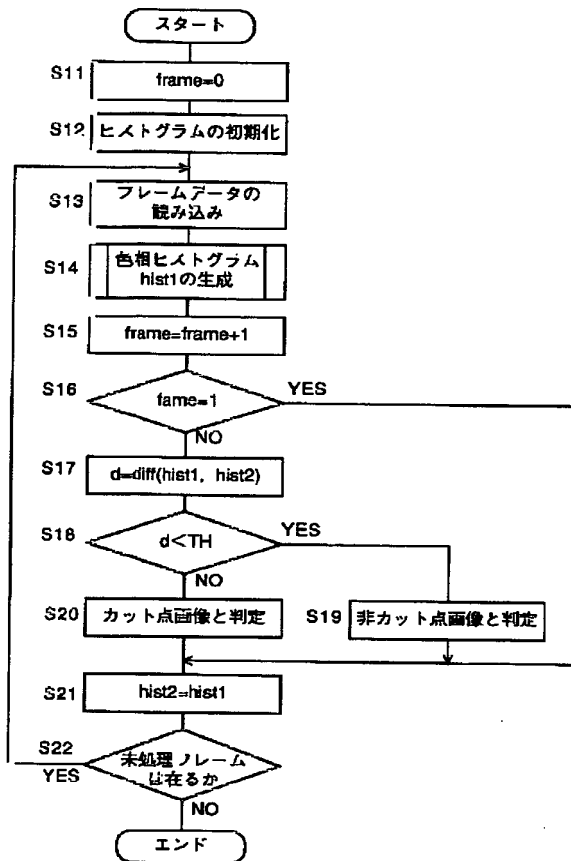
【図2】



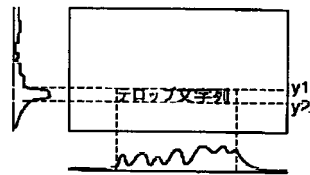
【図4】



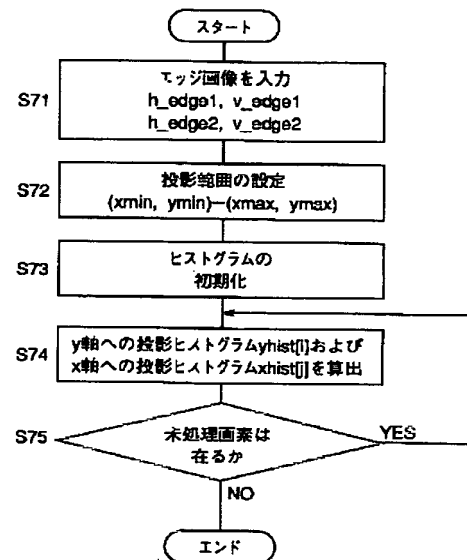
【図3】



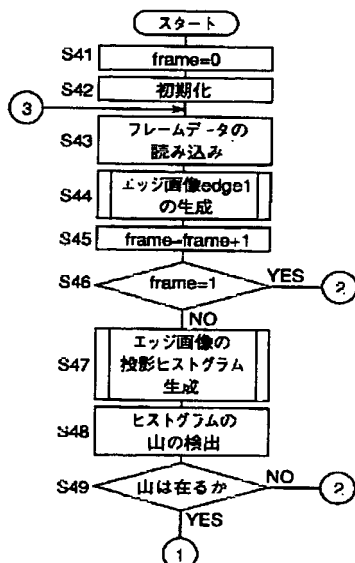
【図6】



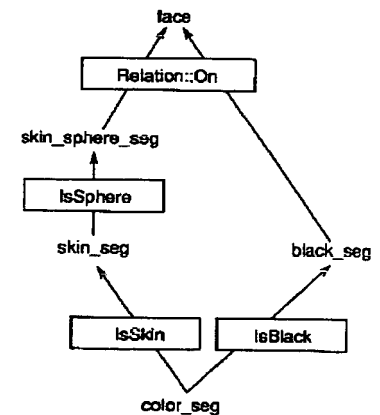
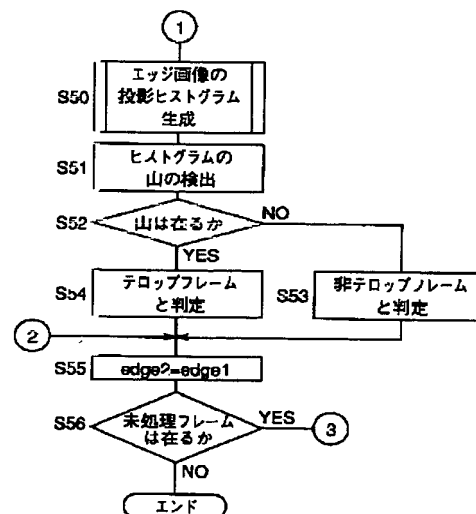
【図8】



【図5】

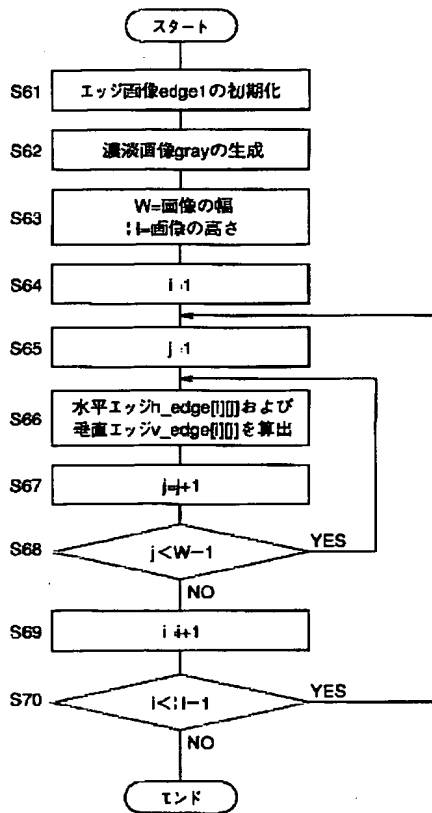


【図10】

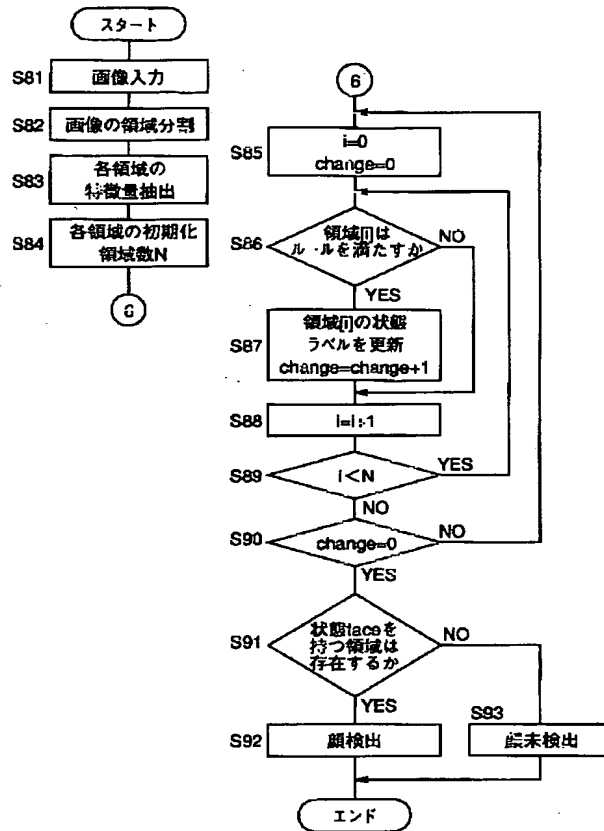




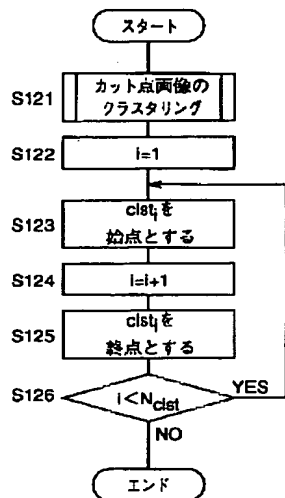
【図7】



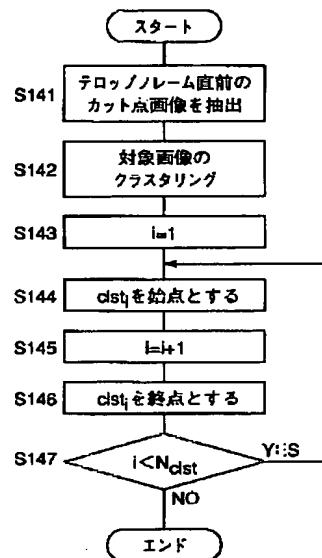
【図9】



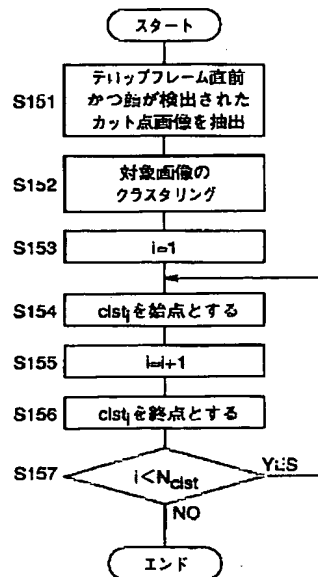
【図12】



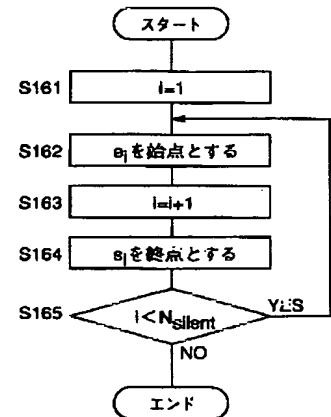
【図14】



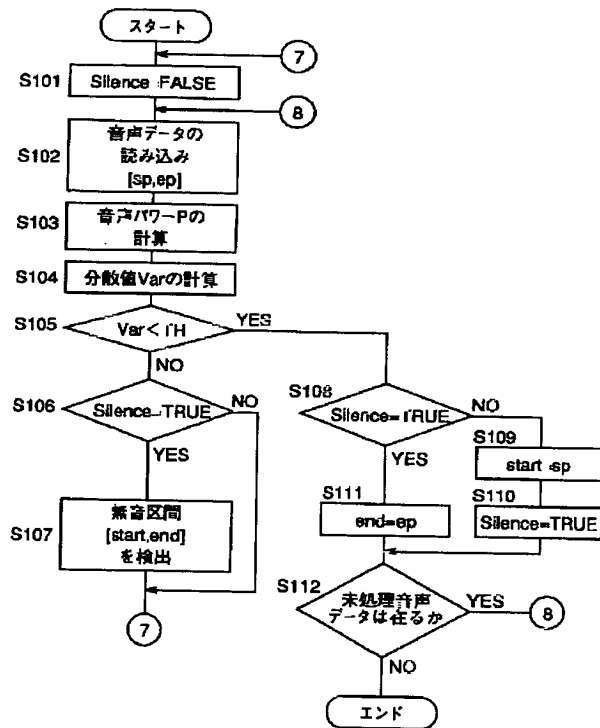
【図15】



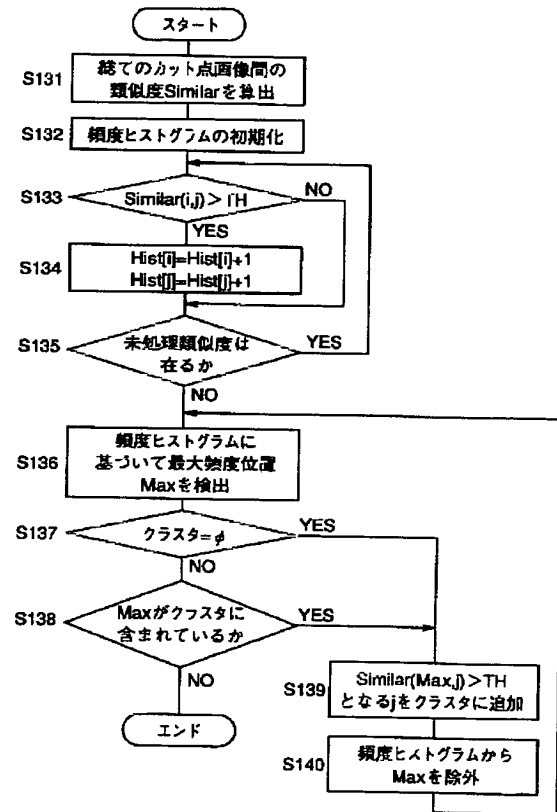
【図16】



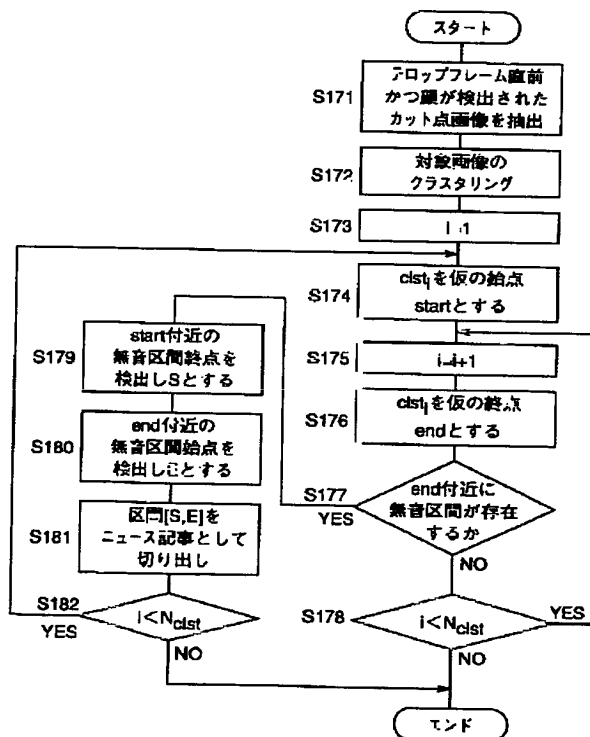
【図11】



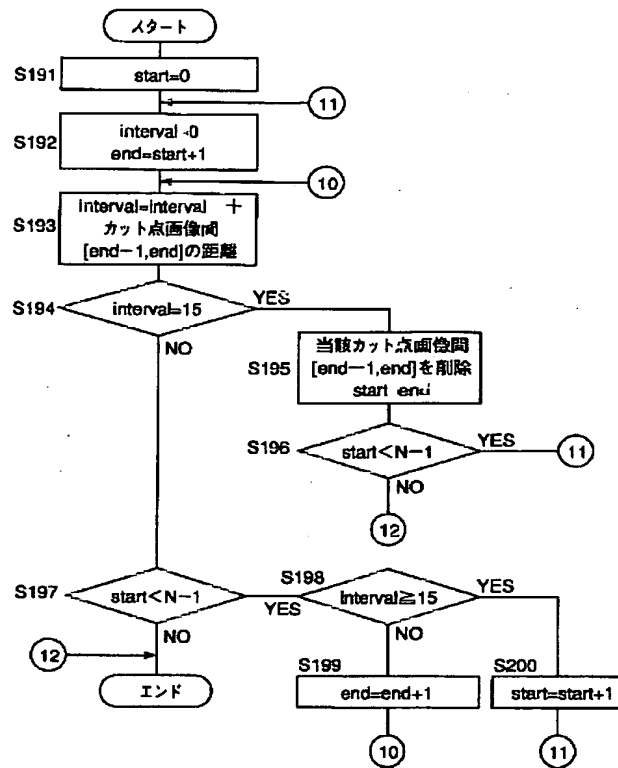
【図13】



【図17】



【 18 】



フロントページの続き

Fターム(参考) 5B050 BA06 BA12 EA08 EA09 EA18  
 5B057 CA01 CA08 CA12 CA16 CB01  
 CB08 CB12 CB16 CE09 DA06  
 DB02 DB06 DB09 DC14 DC16  
 DC25 DC36  
 5L096 AA02 AA06 GA34 GA41  
 9A001 HH28 HH30

**THIS PAGE BLANK (USPTO)**